All the independent variables in multicollinearity are larger than 80 percent. The rules for the project were to drop maximum of 5 variables. If I drop more variables, there will be one or two variables left.

|  |  |  |  |
| --- | --- | --- | --- |
| Variables | R Square for Multicollinearity | Multcollinearity (More than 80%?) | Variance Inflationary Factory |
| U.S. Population (in million) (X1) | 0.9888 | Yes | 89.5560 |
| Personal Saving Rate (X2) | 0.9989 | Yes | 886.6019 |
| Average unemployment rate (X3) | 0.99902 | Yes | 1015.4628 |
| Inflation Rates (X4) | 0.9986 | Yes | 694.4186 |
| Housing Start (in thousands) (X5) | 0.9985 | Yes | 684.9241 |
| National Debt (in Trillions) (X6) | 0.9989 | Yes | 927.3734 |
| Gold (Average Price) (X7) | 0.99903 | Yes | 1030.2016 |
| U.S. Imports (in Billions) (X8) | 0.9965 | Yes | 286.7632 |
| U.S. Imports (in Billions) (X9) | 0.9988 | Yes | 801.4526 |
| The political party of the US President (X10) | 0.99902 | Yes | 1018.5681 |

|  |  |
| --- | --- |
|  | *Coefficients* |
| Intercept | -10136.2126 |
| U.S. Population (in million) (X1) | 62.4612 |
| Personal Savings Rate (X2) | -83.2489 |
| Average unemployment rate (X3) | 23.5805 |
| Inflation Rates (X4) | -58.2224 |
| Housing Start (in thousands) (X5) | -0.4515 |
| National Debt (in Trillions) (X6) | 108.0037 |
| Gold (Average Price) (X7) | -0.1458 |
| U.S. Imports (in Billions) (X8) | 3.6246 |
| U.S. Exports (in Billions) (X9) | -2.1721 |
| The political party of the US President (X10) | -57.6445 |

The numbers in the coefficients, the number are really high and some have negative numbers. In the real world, all the number would be positive in the coefficients. As precision is simply variance, high multicollinearity implies high slope estimator variance. This in turn implies low t-statistics, since the denominator of the standard t-statistic(say H0: beta = 0, then t= beta hat / SE beta hat, where SE is the standard error of the estimator of the slope coefficient - which is reported in any computer output from a program which does regression). These low t-statistics in turn may result in a failure to reject the null hypothesis that some particular slope coefficient is zero, in turn implies that the particular independent variable is not a useful explanatory variable. However, the variable may actually have a lot of explanatory power, and we may simply be fooled into beleiving the variable is irrelevant because we observe low t-statistics which are simply an artifact of the multicollinearity in our regression model. This problem is often signaled when our regression has a high R squared value, but very low slope coefficient t-statistics. On simple remedy is to omit one of the variables that is highly multicollinear, as the informational content of this variable is essentially the same as that of other variable(s), anyways. Another common solution is to difference of log difference the data. This often removes much of the multicollinearity among regressors, particularly since the multicollinearity may have arisen because the regressors were all trending upwards over time, say, which is then the same problem as discussed in the previous note on nonstationarity.

The more common situation of high multicollinearity leads to large variances and covariances, large confidence intervals, and insignificant significance coefficients